

WHAT IS CLAIMED IS:

- 1 1. An apparatus for direct annotation of objects, the apparatus comprising:
 - 2 a display device for displaying one or more images;
 - 3 an audio input device for receiving an audio input; and
 - 4 a direct annotation creation module coupled to the audio input device and the display device, the direct annotation creation module creating an annotation object that associates an input audio signal an image displayed on the display device.
- 1 2. The apparatus of claim 1 further comprising an annotation display module coupled to the direct annotation creation module, the annotation display module generating symbols or text representing the annotation objects.
- 1 3. The apparatus of claim 1 further comprising an annotation audio output module coupled to the direct annotation creation module, the annotation audio output module generating audio output in response to user selection of an annotation symbol representing an annotation object.
- 1 4. The apparatus of claim 1 further comprising:
 - 2 an audio vocabulary storage for storing a plurality of audio signals and
 - 3 corresponding text strings;

4 an audio vocabulary comparison module coupled to the audio input device, the
5 audio vocabulary storage and the direct annotation creation module, the
6 audio vocabulary comparison module receiving audio input and finding a
7 corresponding text string that matches the audio input; and
8 wherein the direct annotation creation module uses text strings found by the audio
9 vocabulary comparison module to create the audio annotation.

1 5. The apparatus of claim 1 further comprising:
2 an audio vocabulary storage for storing a plurality of audio signals and
3 corresponding text strings;
4 a dynamic vocabulary updating module coupled to the audio vocabulary storage
5 and the audio input device, the dynamic vocabulary updating module for
6 displaying an interface to create a new entry in the audio vocabulary
7 storage, the dynamic vocabulary updating module receiving an audio input
8 and a text string and creating the new entry in the audio vocabulary
9 storage.

1 6. The apparatus of claim 1 further comprising a media object cache for storing
2 media and annotation objects.

1 7. An apparatus for direct annotation of objects, the apparatus comprising:
2 a direct annotation creation module coupled to receive an input audio signal and a
3 reference to an image, the direct annotation creation module creating an annotation object
4 that associates a symbol or text with the image; and

5 an annotation display module coupled to the direct annotation creation module,
6 the annotation display module generating the symbol or text representing the annotation
7 object on a display device.

1 8. An apparatus for direct annotation of objects, the apparatus comprising:
2 a direct annotation creation module coupled to receive an input audio signal and a
3 reference to an image, the direct annotation creation module creating an annotation object
4 that associates the input audio signal and the image; and
5 an annotation audio output module coupled to the direct annotation creation
6 module, the annotation audio output module generating audio output in response to user
7 selection of an annotation symbol representing the annotation object.

1 9. An apparatus for direct annotation of objects, the apparatus comprising:
2 a media object storage for storing media and annotation objects; and
3 a direct annotation creation module coupled to receive an input audio signal and a
4 reference to an image, the direct annotation creation module creating an annotation object
5 that associates the input audio signal and the image, the direct annotation creation module
6 storing the audio annotation in the media object storage.

1 10. A method for direct annotation of objects, the method comprising the steps
2 of:
3 displaying an image;
4 receiving audio input;
5 detecting selection of an image; and

6 creating an annotation between the selected image and the audio input.

1 11. The method of claim 10, wherein the step of displaying is performed before
2 or simultaneously with the step of receiving.

1 12. The method of claim 10, wherein the step of receiving is performed before or
2 simultaneously with the step of displaying.

1 13. The method of claim 10, wherein the step of detecting selection includes
2 detecting a portion of the image; and wherein the annotation creates an association
3 between the portion of the image and the audio input.

1 14. The method of claim 10, further comprising the step of displaying a visual
2 notation that the image has an annotation.

1 15. The method of claim 14, wherein the visual notation is text or a symbol.

1 16. The method of claim 10, wherein the step of creating an annotation includes
2 creating an annotation object and storing the annotation object in an object storage.

1 17. The method of claim 10, further comprising the step of recording the audio
2 input received.

1 18. The method of claim 17, wherein the step of creating an annotation includes
2 creating an annotation object and storing the recorded audio input as part of the
3 annotation object.

1 19. The method of claim 10, further comprising the step of comparing the audio
2 input to a vocabulary to produce text.

1 20. The method of claim 19, wherein the step of creating an annotation includes
2 creating an annotation object and storing the text as part of the annotation object.

1 21. The method of claim 10, further comprising the steps of:
2 comparing the audio input to a vocabulary;
3 determining if the audio input has a matching entry in the vocabulary; and
4 storing the entry as part of the annotation object if the audio input has a matching
5 entry in the vocabulary.

1 22. The method of claim 21, further comprising the steps of:
2 determining if the audio input has a close match in the vocabulary;
3 displaying the close matches;
4 receiving input selecting a close match; and
5 storing the selected close match as part of the annotation object if the audio input
6 has a close match in the vocabulary.

1 23. The method of claim 22, further comprising the step of displaying a message
2 that the image has not been annotated if there is neither a matching entry in the
3 vocabulary nor a close match in the vocabulary.

1 24. The method of claim 22, further comprising the following steps if there is
2 neither a matching entry in the vocabulary nor a close match in the vocabulary:
3 receiving text input corresponding to the audio input;
4 updating the vocabulary with a new entry including the audio input and the text
5 input; and
6 wherein the received text is stored as part of the annotation object.

1 25. The method of claim 10, further comprising the steps of:
2 receiving text input corresponding to the audio input;
3 updating the vocabulary with a new entry including the audio input and the text
4 input.

5 26. A method for direct annotation of objects, the method comprising the steps
6 of:
7 displaying an image;
8 receiving audio input;
9 detecting selection of an image;
10 comparing the audio input to a vocabulary to produce text; and
11 creating an annotation between the selected image and the text.

1 27. The method of claim 26, further comprising the step of recording the audio
2 input received.

1 28. The method of claim 27, wherein the step of creating an annotation includes
2 creating an annotation object including a reference to the selected image, the recorded
3 audio input and the text, and storing the annotation object in an object storage.

1 29. The method of claim 26, wherein the step of creating an annotation includes
2 creating an annotation object and storing the text as part of the annotation object.

1 30. The method of claim 26, further comprising the steps of:
2 determining if the audio input has a matching entry in the vocabulary; and
3 storing the entry as part of the annotation object if the audio input has a matching
4 entry in the vocabulary.

1 30. The method of claim 29, further comprising the steps of:
2 determining if the audio input has a close match in the vocabulary;
3 displaying the close matches;
4 receiving input selecting a close match; and
5 storing the selected close match as part of the annotation object if the audio input
6 has a close match in the vocabulary.

1 31. The method of claim 30, further comprising the step of displaying a message
2 that the image has not been annotated if there is neither a matching entry in the
3 vocabulary nor a close match in the vocabulary.

1 32. The method of claim 30, further comprising the following steps if there is
2 neither a matching entry in the vocabulary nor a close match in the vocabulary:
3 receiving text input corresponding to the audio input;
4 updating the vocabulary with a new entry including the audio input and the text
5 input; and
6 wherein the received text is stored as part of the annotation object.

1 33. The method of claim 26, further comprising the steps of:
2 receiving text input corresponding to the audio input;
3 updating the vocabulary with a new entry including the audio input and the text
4 input.

1 34. A method for displaying objects with annotations, the method comprising the
2 steps of:
3 retrieving an image;
4 displaying the image with a visual notation that an annotation exist;
5 receiving user selection of an image; and
6 outputting a notation associated with the selected image.

1 35. The method of claim 34, wherein the annotation is text and the step of
2 outputting is displaying the text proximate an image that it annotates.

1 36. The method of claim 34, wherein the annotation is an audio signal and the
2 step of outputting is playing the audio signal.

1 37. The method of claim 34, further comprising the steps of:
2 determining whether the annotation includes text;
3 retrieving a text annotation for the selected image; and
4 displaying the retrieved text with the image.

1 38. The method of claim 34, further comprising the steps of:
2 determining whether the annotation includes an audio signal;
3 retrieving a audio signal for the selected image; and
4 wherein the step of outputting is playing the audio signal.

1 39. A method for retrieving images, the method comprising the steps of:
2 receiving audio input;
3 determining annotation objects that reference a close match to the audio input;
4 retrieving the images that are referenced by the determined annotation objects;
5 and
6 displaying the retrieved images.

1 40. The method of claim 39, wherein the step of determining annotation objects
2 further comprising the steps of:
3 comparing the audio input to an audio signal reference by an annotation object;
4 and
5 determining a close match between the audio input to the audio signal reference
6 by an annotation object if a probability metric is greater than an a
7 threshold of 80%.

1 41. The method of claim 39, wherein the step of determining annotation objects
2 further comprising the steps of:
3 determining the annotation objects for a plurality of images;
4 for each annotation object, comparing the audio input to an audio signal reference
5 by an annotation object; and
6 determining a close match between the audio input to the audio signal reference
7 by an annotation object if a probability metric is greater than an a
8 threshold of 80%.